AN ABSTRACT OF THE THESIS OF

Ali Ahmed Rahuma for the degree of Master of Science

Agricultural and in Resource Economics presented on November 17, 1979

Title: APPLICATION AND EVALUATION OF RIDGE REGRESSION TO SELECTED EMPIRICAL ECONOMIC MODELS

Abstract approved Redacted for privacy

W. G. Brown

When the assumption of independency between the independent variables in a regression equation is violated, and in spite of Ordinary Least Square (OLS) procedure yielding estimates that are best linear unbiased (BLUE), the presence of multicollinearity can have severe effects on the estimation of the coefficients and on variable selection techniques.

In an attempt to cope with the problem of multicollinearity in a regression analysis, a number of techniques have been proposed. Among these techniques, the deletion of variables most affected by multicollinearity, the use of prior information models and the use of ridge regression are described and commented upon. Advantages and disadvantages are stated in this study. The sources of multicollinearity, its harmful effects and several methods of its detection are presented and discussed.

The purpose of this study is to evaluate ridge regression as a
possible method in obtaining stable and reliable estimates when multi-
collinearity effects are present in economic models. In order to do
that, four different economic models suffering from various degrees
of multicollinearity have been chosen, and three versions of ridge
regression estimators have been applied to these economic models.

Results of these studies indicate that the three ordinary ridge
estimators, one proposed by Horel, Kennard and Baldwin ($K_A$), the
second proposed by Lawless and Wang ($K_B$) and the third introduced
by Dempster, Schatzoff and Wermuth (RIDGM) appear to be effective
in obtaining stable and reliable estimates when multicollinearity is a
serious problem in a regression analysis.

Although the estimates obtained by ridge regression in this
study are biased, they appear to have desirable characteristics,
are all stable and have the correct sign, and lack the symptoms of
nonsense regression that was observed when OLS was used to
estimate these models with highly multicollinear data.
Application and Evaluation of Ridge Regression
to Selected Empirical Economic Models

by

Ali Ahmed Rahuma

A THESIS
submitted to
Oregon State University

in partial fulfillment of
the requirements for the
degree of

Master of Science
June 1978
APPROVED:

Redacted for privacy
Professor of Agricultural and Resource Economics
in charge of major

Redacted for privacy
Head of Department of Agricultural and Resource Economics

Redacted for privacy
Dean of Graduate School

Date thesis presented November 17, 1977

Typed by Susie Avery for Ali Ahmed Rahuma
ACKNOWLEDGMENT

I wish to express my appreciation to my major professor Dr. W. Brown for his assistance and guidance throughout the course of this study, and to other members of my graduate committee, Dr. Roger Petersen, Dr. Richard Johnston, and Dr. Robert McMahon.

I especially acknowledge Dr. G. Blanch for his guidance and advice throughout my academic study at Oregon State University.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>1</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td></td>
</tr>
<tr>
<td>Multicollinearity</td>
<td>3</td>
</tr>
<tr>
<td>Definition</td>
<td>4</td>
</tr>
<tr>
<td>Sources of Multicollinearity</td>
<td>4</td>
</tr>
<tr>
<td>Detection of Multicollinearity</td>
<td>7</td>
</tr>
<tr>
<td>Consequences of Multicollinearity</td>
<td>10</td>
</tr>
<tr>
<td>II</td>
<td>14</td>
</tr>
<tr>
<td>SUGGESTED SOLUTIONS TO MULTICOLLINEARITY</td>
<td></td>
</tr>
<tr>
<td>Deletion of Variables</td>
<td>14</td>
</tr>
<tr>
<td>The Use of Prior Information</td>
<td>18</td>
</tr>
<tr>
<td>The Theil-Goldberger Mixed Model</td>
<td>19</td>
</tr>
<tr>
<td>Prior Information in the Form of Inequality Restraints</td>
<td>21</td>
</tr>
<tr>
<td>Use of Ridge Regression</td>
<td>23</td>
</tr>
<tr>
<td>III</td>
<td>26</td>
</tr>
<tr>
<td>BIASED LINEAR ESTIMATIONS</td>
<td></td>
</tr>
<tr>
<td>Stein Estimators</td>
<td>26</td>
</tr>
<tr>
<td>Ridge Regression</td>
<td>28</td>
</tr>
<tr>
<td>IV</td>
<td>31</td>
</tr>
<tr>
<td>APPLICATION OF SOME RIDGE REGRESSION MODELS FOR THE ESTIMATION OF ECONOMIC RELATIONSHIPS</td>
<td></td>
</tr>
<tr>
<td>Nature of Economic Data</td>
<td>31</td>
</tr>
<tr>
<td>Ridge Regression in Estimating the Marginal Value Productivity of Irrigation Water</td>
<td>32</td>
</tr>
<tr>
<td>A Production Function Analysis of Water Resource Productivity in Pacific Northwest Agriculture</td>
<td>40</td>
</tr>
<tr>
<td>Consequences of Multicollinearity in the Model</td>
<td>44</td>
</tr>
<tr>
<td>Salmon Steelhead Fishing Demand Relationship</td>
<td>50</td>
</tr>
<tr>
<td>Estimating the Demand Function for Commer- cially Caught Salmon in the Columbia River</td>
<td>57</td>
</tr>
<tr>
<td>V</td>
<td>63</td>
</tr>
<tr>
<td>SUMMARY AND CONCLUSIONS</td>
<td></td>
</tr>
<tr>
<td>Limitations and Additional Research Needed</td>
<td>65</td>
</tr>
<tr>
<td>BIBLIOGRAPHY</td>
<td>67</td>
</tr>
<tr>
<td>Table</td>
<td>Description</td>
</tr>
<tr>
<td>-------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>1</td>
<td>Estimated values for regression of county values on all farm products sold as a Cobb-Douglas function of inputs, OLS estimates for 25 Central Pacific counties, 1954</td>
</tr>
<tr>
<td>2</td>
<td>Simple correlation coefficient for regression of county values of all farm products sold as a Cobb-Douglas function of inputs</td>
</tr>
<tr>
<td>3</td>
<td>Estimates of the betas, variances and estimated MSE, OLS vs. $K_A$, $K_B$ and $K_M$ for regression of county values of all farm products sold as a Cobb-Douglas function of inputs</td>
</tr>
<tr>
<td>4</td>
<td>Variances and estimated MSE of coefficients, OLS vs. $K_A$, $K_B$ and $K_M$ (OLS = 100) for regression of county values of all farm products sold as a Cobb-Douglas function of inputs</td>
</tr>
<tr>
<td>5</td>
<td>Ordinary least square estimates, standard errors, means, and t-values for analysis of water resource productivity in Pacific Northwest (Area C) 1964</td>
</tr>
<tr>
<td>6</td>
<td>Water resource productivity in Pacific Northwest agriculture (Area C) simple correlation</td>
</tr>
<tr>
<td>7</td>
<td>Variance inflation factors on the diagonal elements of the inverted correlation matrix (XX) Pacific Northwest water resource productivity data (Area C) 1964</td>
</tr>
<tr>
<td>8</td>
<td>Estimated coefficients, and standard errors, OLS, Theil-Goldberger mixed model and ridge regression for the analysis of water resource productivity in Pacific Northwest (Area C) 1964</td>
</tr>
<tr>
<td>9</td>
<td>Simple correlation coefficients for salmon steelhead fishing demand relationship</td>
</tr>
</tbody>
</table>
(List of Tables Continued)

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>54</td>
</tr>
<tr>
<td>11</td>
<td>55</td>
</tr>
<tr>
<td>12</td>
<td>59</td>
</tr>
<tr>
<td>13</td>
<td>60</td>
</tr>
<tr>
<td>14</td>
<td>61</td>
</tr>
</tbody>
</table>

Salmon steelhead demand relationship coefficients, variances and estimated MSE, OLS vs. ridge regression

Salmon steelhead demand relationship coefficients, variances and estimated MSE, OLS vs. ridge regression (with restriction on the coefficients)

Coefficients, means, and standard deviations for commercially caught salmon in Columbia River demand relationship

Simple correlation coefficients between variables for commercially caught salmon in Columbia River

Demand relationship for the commercially caught salmon in Columbia River, coefficients, variances, and estimated mean square errors, OLS vs. $K_A$, $K_B$ and $K_M$ estimators
APPLICATION AND EVALUATION OF RIDGE REGRESSION TO SELECTED EMPIRICAL ECONOMIC MODELS

I. INTRODUCTION

A serious problem that can occur in a regression analysis is the presence of multicollinearity among the independent variables in a regression equation.

The problem is unavoidable in most economic relationships (1), due to the nature of the economic data available. These data are usually aggregated over time, or over geographical location (cities and counties). In any case, a researcher is usually interested in cases of dependency or interrelationships among variables rather than in states of independency.

When the assumption of independency among explanatory variables in a regression equation is violated, and in spite of the Ordinary Least Square (OLS) procedure yielding estimates that are best linear unbiased estimates (BLUE), the presence of multicollinearity will destroy the quality of these estimates. This happens because the variances of the estimates are unreasonably large. It also results in rejecting variables for lack of statistical significance, or estimates of the wrong sign are obtained.

1/ Vectors $\mathbf{x}_1', \mathbf{x}_2', \ldots, \mathbf{x}_p$ are linearly dependent if there exist non-zero constants, $a_1, a_2, \ldots, a_p$ such that $\sum_{j=1}^{p} a_j x_j = 0$.
Many researchers and economists concern themselves with the problem of multicollinearity in order to improve the situation. There are two ways that multicollinearity can be handled in a regression analysis. One way is to delete the variable most affected in an attempt to obtain more accurate and reliable estimates. There are some advantages and many disadvantages to this technique. Few economists have applied the ridge regression techniques to economic data (2) since most of the ridge studies are based on non-economic data.

The purpose of this study is to apply some ridge regression models which are shown to have promising results for improving estimates with multicollinearity present (10). These ridge regression models are different only in the choice of $K$, but they are taken from Hoerl, Baldwin and Kennard (11), Lawless and Wang (15), and Dempster, Shatzoff and Wermuth (5).

The plan of this paper is as follows: multicollinearity will be discussed in chapter one; in chapter two, some suggested remedies for multicollinearity will be presented; in chapter three, some models of ridge regression will be discussed; and chapter four will apply the ridge regression models to four different economic models discussed in the literature (2).

The use of alternatives to variable deletion is both possible, and reliable in some cases, and economists are coming to accept the technique more readily. It is being discovered increasingly how to
break the multicollinearity deadlock. Different techniques are dis-
cussed in past studies by Theil (19), Goldberger (8), and Brown (2),
by which prior information can be incorporated into regression analysis.

Ridge regression, introduced by Hoerl and Kennard (1968), is
among the techniques benefitting from prior information. Since then,
economists and statisticians have been developing the technique:
Lawless and Wang (16), Dempster, Schatzoff and Wermuth (5). When
multicollinearity is present, ridge regression techniques result in
estimates of the coefficients that are biased, but have smaller vari-
ances than that of Ordinary Least Square (OLS), as a result of intro-
ducing some bias into the (X’X) data matrix. Hoerl and Kennard (10),
Lawless and Wang (6), Dempster (5), Brown (1) and others have con-
sidered ridge regression as a possible remedy for improving esti-
mates when multicollinearity is a problem in regression analysis.

Multicollinearity

Multicollinearity is a situation in which one cannot single out
or separate the effects of two or more explanatory variables, because
they tend to move together in sample data. It is the degree of inter-
dependency between two or more theoretically independent variables.

Because of the nature of the economic and social relationships
whose data tend to move together, we measure parameters of dependency
and not independency. Multicollinearity is then an unavoidable case.
Multicollinearity is defined by Farrar and Glauber (7) as "a departure from Orthogonality in an independent variable set." With this definition in hand, two things can now be separated:

(1) The nature of the problem of multicollinearity: the lack of independence between explanatory variables in the model.

(2) The effects and locations of multicollinearity: to what degree variables are affected by multicollinearity.

Sources of Multicollinearity

There are many sources of multicollinearity. Only four are well known in the literature (see 1, 2, 6, 17, 19 and 20).

(1) Aggregation of data.

(2) Specification of the model.

(3) Sampling technique.

(4) Some physical limitations on the model.

Each of these sources presents a problem in the detection of multicollinearity and its effects.

Aggregation of data: a common practice, resulting from the bulkiness of the information and the difficulty of handling each individual factor of this information. Economists and researchers very often depend in their analyses on aggregated data based on
available sources, such as an Agricultural or Commercial census.
Aggregation will result in loss of information, and it will create
dependency between economic variables with multicollinearity arising
as a result.

**Specification of the model:** in many cases, especially in medical
and industrial sections, a model may be specified in such a way that it
becomes over-defined, a situation in which one has more variables in
the model than observations. There is then a serious problem.

**Sampling technique:** this technique results in multicollinearity
when the investigator only samples a subspace of the experiment space.
An omission of important information is caused, often leading to
dependency between one or more of the explanatory variables in the
model.

**Physical limitation of the model:** the experimenter is confronted
with some limitation in the data such as in the presence of fixed pro-
portionality, or simply that some information is not available.

It is important to recognize the different sources of multi-
collinearity in an attempt to remove it or to assist in the discovery of
some workable solutions to the problem.

In estimating parameters of an economic model in the form

\[ y = x\beta + u \]

where \( y \) is \((n \times 1)\) vector of dependent variables, \( x \) is \((n \times p)\) matrix
of theoretically independent explanatory variables, \( \beta \) is \((p \times 1)\) vector
of coefficients and \( u \) is \((n \times 1)\) vector of terms, where \( u \sim N(0, 2) \). OLS estimates

\[
\hat{\beta} = (xx)^{-1} x'y
\]

and

\[
\text{Var-Cov} (\hat{\beta}) = \sigma (xx)^{-1}.
\]

Two extreme cases may be recognized.

(1) The orthogonal case where \((xx)\) is an orthogonal matrix with zeros in all the off diagonal elements because of perfect independence between each explanatory variable. Since

\[
\begin{align*}
    r_{ij} &= \begin{cases} 
0 & \text{if } i \neq j \\
1 & \text{if } i = j
\end{cases}
\end{align*}
\]

the OLS estimates are BLUE with minimum bias. \( V(\hat{\beta}_i) \) is small and \( \hat{\beta}_i \) is not sensitive to change in data or the specification of the model.

(2) The most drastic cases of multicollinearity where \((xx)^{-1}\) does not exist, or is hard to obtain using the existing computational techniques. This is a case of perfect multicollinearity between two or more explanatory variables: the determinant of \((xx)\) is zero and \((x'x)\) is not invertable. OLS estimates \( \hat{\beta} \) are not possible and the variance \( V(\hat{\beta}_i) \) explodes at the limit.

Between these two extremes on the multicollinearity scale, there are degrees of multicollinearity. 'Harmful multicollinearity,' a term defined as "one that causes a wrong sign or symptoms of nonsense..."
regression" (7), concerns researchers the most, because of difficulties it creates in estimation and model specification. The problem is of estimation because the coefficients are sensitive to change in data, in specification. Wrong signs of some coefficients may mislead the researcher to a wrong decision of deleting these variables.

Detection of Multicollinearity

(1) A simple measure for the detection of multicollinearity is through the elements of the correlation matrix \( (x'x) \). \( r_{ij}^{2/} \) is near unity when there is perfect collinearity and it is equal to zero when the two variables \( x_i \) and \( x_j \) are not collinear. Unfortunately, this measure is good only in the two factor case, and it is not useful when more than two factors are in the model.

(2) Another method which is widely practiced for the detection of multicollinearity is by regressing \( y \) on all the explanatory variables in the model except \( x_j \). Compare \( R^2_{3/} \) with \( R_j^2 \) where \( x_j \) is excluded from the model. \( R_j^2 \) will be close to \( R_j^2 \) in cases of severe multicollinearity. Unfortunately this technique is not very effective in the detection of multicollinearity, because it will reflect

\[
2/ r_{ij}^{2/} = \frac{\sum x_i x_j}{\sqrt{x_i^2} \sqrt{x_j^2}} \quad \text{where } x_i \text{ and } x_j \text{ are in deviation form from their respective means.}
\]

\[
3/ R^2 = \text{regression sum of squares/total sum of squares.}
\]
only the importance of $x_j$ as a predictor and not necessarily that $x_j$ is collinear with other variables in the model.

(3) Another simple measure for the detection of multicollinearity when $(xx)$ is standardized, is that

$$\text{Det}(x'x) = |x'x| = 0$$

when exact relationships are present between two or more columns of $(x'x)$ and $|x'x| = 1$ indicates orthogonal columns of the data matrix.

(4) Different methods for the detection of multicollinearity are suggested by Webster, Grunst and Mason (17). The idea of this test is to utilize the smallest Eigenvalues ($\lambda_{\text{min}}$) and the corresponding Eigenvector $V$.

The closer $\lambda_{\text{i}}$ is to zero, the stronger the linear relationship among the columns or rows of $(x)$ matrix.

The following is a more recent approach for the detection of multicollinearity.

(5) To decide how to detect multicollinearity and how to determine its location:

As before assume $y = x\beta + u$ is the general linear model with

$$\hat{\beta} = (x'x)^{-1}x'y$$

and

$$\text{var - cov}(\hat{\beta}) = \sigma^2(x'x)^{-1}.$$ 

Following Glauber and Farrar (7)

$(x'x)$ is the correlation matrix
$R = (X'X) = \begin{bmatrix}
    r_{11} & r_{12} & \cdots & r_{1p} \\
    r_{21} & r_{22} & \cdots & r_{2p} \\
    \vdots & \vdots & \ddots & \vdots \\
    r_{p1} & r_{p2} & \cdots & r_{pp}
\end{bmatrix}$

\[
    r_{ij} = \frac{\sum_{i=1}^{p} x_i x_j}{\sqrt{x_i^2} \sqrt{x_j^2}}
\]

$(x_i)$ is measured in a deviation form from its mean. The "variance inflation factor" or the diagonal elements of the inverted correlation matrix $(x'x)$ is a good measure of the degree and location of multicollinearity. Let it be

\[
    C^{ii} = \frac{|(x'x)_{ii}|}{|(x'x)|}
\]

where $(x'x)_{ii}$ denotes the correlation matrix excluding the $i^{th}$ variable $x_i$ and $(x'x)$ is the whole correlation matrix

$1 \leq C^{ii} \leq \infty$

$C^{ii}$ the diagonal element or "variance inflation factor", equal to unity, when there is no multicollinearity. Since in this case, $(x'x)_{ii}$ will be the same as $(x'x)$ and the ratio of

\[
    \frac{|(x'x)_{ii}|}{|(x'x)|} = 1.0
\]
When $(x'x)$ is singular, and there is perfect collinearity between two or more explanatory variables in the variable set.

The bigger $C^{ii}$ is, the more serious the problem of multicollinearity because $V(\hat{\beta}_i)$ will be inflated by this factor $(C^{ii})$ and the usual $(t)$ test will reject more variables $(x_i)$ as statistically insignificant due to the small $t$ values. As $V(\hat{\beta}_i)$ increases $t$ becomes small and may not reflect the true importance of $x_i$ as an explanatory variable.

As seen from this test, the partial correlation matrix $R$ is not a sufficient test for the detection of multicollinearity, because the latter is a question of degree relative to the overall level of association.

**Consequences of Multicollinearity**

(1) Because of the lack of independence among the variables in the set of explanatory variables in the model, the precision of the estimates falls due to the difficulty in separating individual variable effects from other variable effects. This difficulty will create large estimate errors and the variance of $\hat{\beta}_i$, $V(\hat{\beta}_i)$ will be unreasonably large. The following example will illustrate this point.

Assume the following: A model with two explanatory variables $x_1, x_2$

\[ t = \frac{\hat{\beta}_i - E(\hat{\beta}_i)}{S \cdot E(\hat{\beta}_i)} \]
\[
(x'x) = \begin{bmatrix}
1 & r_{12} \\
r_{12} & 1
\end{bmatrix}
\]

and

\[
V(\hat{\beta}_i) = \sigma_u^2 (x'x)^{-1} = \frac{\sigma_u^2}{1-r_{12}^2} \begin{bmatrix}
1 & -r_{12} \\
r_{12} & 1
\end{bmatrix}
\]

in a situation near orthogonality with \( r_{12} = 0.20 \)

\[
V(\hat{\beta}_i) = \sigma_u^2 \begin{bmatrix}
1.04 & -.208 \\
-.208 & 1.04
\end{bmatrix}
\]

Now assume another case where \( r_{12} = 0.90 \). Due to multicollinearity, \( V(\hat{\beta}_i) \) will increase to

\[
V(\hat{\beta}_i) = \sigma_u^2 \begin{bmatrix}
10 & -9 \\
-9 & 10
\end{bmatrix}
\]

i.e. \( V(\hat{\beta}_i) \) increased by a multiple of about (10).

2) Estimates obtained by OLS will be very sensitive to change in data, i.e., an addition of more observations may result for instance in a change in variance of the estimates. This can be seen by

\[
V(\hat{\beta}_i) = \sigma_u^2 (x'x)^{-1}
\]

in addition to the effect of multicollinearity on \( (x'x)^{-1} \), \( \sigma_u^2 \) also will be affected,
\[ \sigma_u^2 = \frac{1 - R^2}{N - p} \]

A change in data, as in the case of adding more observations, \( N \) will be increased and \( \sigma_u^2 \) will be more stable.

(3) Estimates obtained by OLS when multicollinearity is present will be sensitive to model specification. For example, let the model be specified to include another explanatory variable \( x_3 \), and let one assume (for simplicity) that \( r_{12} = r_{13} = r_{23} = r \). Then

\[
(x'x) = \begin{bmatrix} 1 & r & r \\ r & 1 & r \\ r & r & 1 \end{bmatrix}
\]

\[
V(\hat{\beta}) = \sigma_u^2 (x'x)^{-1} = \frac{\sigma_u^2}{2r^3 - 3r^2 + 1} \begin{bmatrix} 1 - r^2 & r^2 - r & r^2 - r \\ r^2 - r & 1 - r^2 & r^2 - r \\ r^2 - r & r^2 - r & 1 - r^2 \end{bmatrix}
\]

\[
V(\hat{\beta}_1) \text{ for example } = V(\beta_1) = \sigma_u^2 \frac{1 - r^2}{2r^3 - 3r^2 + 1} \text{ instead of }
\]

\[
V(\beta_1) = \sigma_u^2 \left( \frac{1}{1 - r^2} \right) \text{ when } p = 2.
\]

When \( r = 0.9 \)

\[
V(\hat{\beta}_1) = \sigma^2 (5.263) \text{ for } p = 2
\]

\[
V(\hat{\beta}_1) = \sigma^2 (6.786) \text{ for } p = 3,
\]

i.e. \( V(\hat{\beta}_1) \) increased by about 29\% when changing the specification.
of the model to include another explanatory variable. Although

\[
\begin{pmatrix}
  r_{11} & r_{12} & \cdots & r_{1p} \\
  r_{21} & r_{22} & \cdots & r_{2p} \\
  \vdots & \vdots & \ddots & \vdots \\
  r_{p1} & r_{12} & \cdots & r_{pp}
\end{pmatrix}
\]

\[
(x'x) = R = \begin{pmatrix}
  r_{11} & r_{12} & \cdots & r_{1p} \\
  r_{21} & r_{22} & \cdots & r_{2p} \\
  \vdots & \vdots & \ddots & \vdots \\
  r_{p1} & r_{12} & \cdots & r_{pp}
\end{pmatrix}
\]

\( r_{ij} \) may be small, but \((x'x)^{-1}\) may not exist and \(|(x'x)| = 0 \).

Glauber writes, "As viewed here, multicollinearity is a property of the independent variable set alone. No account whatever is taken of the extent, or even the existence, of dependence between \(y\) and \(x\)."
II. SUGGESTED SOLUTIONS TO MULTICOLLINEARITY

In this section, some suggested solutions for the removal of multicollinearity will be discussed. Some of these options are:

1. Deletion of variables proven to be highly affected by multicollinearity.

2. Use of prior information to estimate regression coefficients.

3. The Theil-Goldberger mixed model.

4. Use of Ridge Regression.

Deletion of Variables

As a consequence of multicollinearity, some researchers may choose to drop one or more explanatory variables and specify the model in terms of the remaining variables that were not affected by multicollinearity. This practice is useful either to obtain OLS estimates which are otherwise not possible because of the singularity of \((x'x)\), or to obtain more precision in the estimates by reducing the error created by interdependency between a set of explanatory variables in the model, i.e., to obtain small \(\hat{\beta}_1\). Although this solution has been popular, many econometricians are aware of the bias it introduces to the model, or what is known as "specification errors."

\(5/\) A matrix \((x)\) is singular if the \(\text{Det}(x) = |x| = 0\)
Since many types of error can be categorised as specification error, only one type that results from the omission of one or more relevant variable, will be discussed.

Let a true linear model be estimated in the form

\[ y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u \]

Where \( y \) is a dependent variable, \( x_1, x_2, \) and \( x_3 \) are theoretically independent variables. \( u \) is an error term. (All variables are measured in a deviation from their respective means.)

Let one assume that instead of estimating the whole model, the researcher chooses to drop \( x_2 \) and \( x_3 \) and to fit the model in terms of \( x_1 \) only.

\[ y = \hat{\beta}_1 x_1 + u \]

The ordinary least square estimates of the coefficient of \( x \) will be

\[ \hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_{1i} y_i}{\sum x_{1i}^2} \]

Contrary to the fact that this estimate \( \hat{\beta}_1 \) ignores completely the effect of \( x_2 \) in the estimate of the coefficient of \( x_1 \), the true estimate of \( \beta_1 \), as expected under the full model will be

\[ E(\hat{\beta}_1) = \beta_1 + \beta_2 \frac{\sum x_{1i} x_2}{\sum x_{1i}^2} + \hat{\beta}_3 \frac{\sum x_{1i} x_3}{\sum x_{1i}^2} \]
if \( y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \) is the true model and not the fitted
\[
y = \beta_1 x_1 + u
\]
\((\hat{\beta}_1)\) estimates will not be \( \frac{\Sigma x_1 y}{\Sigma x_1^2} \) but a bias is introduced.

This bias is in the form of
\[
\hat{\beta}_2 \frac{\Sigma x_1 x_2}{\Sigma x_1^2} + \hat{\beta}_3 \frac{\Sigma x_1 x_3}{\Sigma x_1^2}
\]
and will depend on

1. the coefficients of the dropped explanatory variables \((\hat{\beta}_2)\) and \((\hat{\beta}_3)\),
2. the quantities:
\[
\frac{\Sigma x_1 x_2}{\Sigma x_1^2} \quad \text{and} \quad \frac{\Sigma x_1 x_3}{\Sigma x_1^2}.
\]
The bias will not be zero unless \( \hat{\beta}_2 \) and \( \hat{\beta}_3 \) are zeros because \( \Sigma x_1 x_2 \) or \( \Sigma x_1 x_3 \) will not be zero, and this is also the case for \( \Sigma x_1^2 \).

This case can be generalised to \( K \) independent variables, as in the following
\[
y = \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + u +
\]
\[
E(\beta_1) = \beta_1 + \beta_2 \frac{\Sigma x_1 x_2}{\Sigma x_1^2} + \beta_3 \frac{\Sigma x_1 x_3}{\Sigma x_1^2} + \ldots + \beta_k \frac{\Sigma x_1 x_k}{\Sigma x_1^2}.
\]
So whenever an important variable is deleted, OLS estimates will be
biased and the degree of the bias will depend upon the relative share of the deleted variable in explaining the regression relationships, i.e., on \( \hat{\beta}_1, \hat{\beta}_3, \ldots \) etc. For more discussion on the general case see Johnston (14).

There are only trivial cases where no bias is introduced from deleting the variable. In such a case, the deleted variables are irrelevant, i.e., where \( \hat{\beta}_1 \) of the \( x_i \) omitted, is zero. But this case is unlikely because if \( x_i \) is irrelevant, it will not be in the specification of the model in the first place.

Although OLS is a biased procedure in cases of deleted variables, it still produces a minimum variance

\[
V(\beta_i) + \frac{\sum x_i^2}{\sum x_i^2}
\]

But the bias created in the estimates can sometimes outweigh the reduction in variance of \( \hat{\beta}_i \), \( V(\hat{\beta}_i) \)

if

\[
\text{MSE}(\beta_i) = \text{var}(\hat{\beta}_i) + \text{Bias}^2(\hat{\beta}_i)
\]

\( \text{MSE}(\hat{\beta}_i) \) of the full model < \( \text{MSE}(\hat{\beta}_i) \) with incomplete model due to the bias factor introduced to the system.

Roa and Miller (18) pointed out that researchers should not be

\[
\text{Bias}(\hat{\beta}_i) = E(\hat{\beta}_i - \beta)
\]
discouraged by the reduction in variances of the estimates they obtain from deleting variables because this reduction should be viewed in terms of MSE ($\hat{\beta}_i$) and not just in terms of the $V(\hat{\beta}_i)$.

The Use of Prior Information

Most econometricians and applied researchers agree that additional information is needed to lessen difficulties of multicollinearity. Various methods of adding prior information into regression analysis to help account for multicollinearity may be used. Prior information could be in the form of:

(1) Exact prior information where the investigator knows exactly some or all of the coefficients and their standard errors, either from previous studies or on a theoretical basis. This information can be incorporated into the regression analysis to ameliorate the difficulties of multicollinearity.

(2) Subjective probability estimates of the coefficients and their variances may be used similarly to solve multicollinearity. Information in the form of inequality is based on the researcher's knowledge about the coefficients from previous experience. It is in the form of probabilities and inequalities such as

$$0 \leq \beta_i \leq 1.0$$

All this information can be used to help solve multicollinearity problems.
The Theil-Goldberger Mixed Model

Statistical prior information can be incorporated into a regression analysis using a method suggested by Theil and Goldberger (20) and summarized as follows: Let the usual linear model expressed as \( y = x\beta + u \) where \( y \) is \((n \times 1)\) vector of the dependent variable, \( x \) is \((n \times p)\) matrix of independent variables, \( \beta \) is \((n \times 1)\) vector of parameters and \( u \) is \((n \times 1)\) vector of disturbances.

\[
E(u) = 0, \quad E(u'u) = \sigma^2 I_n,
\]

\[
\hat{\beta} = (x'x)^{-1} x'y, \quad \text{var}(\hat{\beta}_1) = \sigma^2 (x'x)^{-1}
\]

and the usual assumptions apply.

Let the statistical information given in the form

\[
r = R\beta + V
\]

apply, where \( r \) is a known \((k \times 1)\) vector of estimates for \( R\beta \), \( R \) is a \((k \times r)\) matrix of known fixed elements determining which parameters have prior information, and how this information is weighted, and \( \beta \) is \((r \times 1)\) vector of fixed unknown parameters and \( V \) is \((k \times 1)\) vector of prior information errors.

Also assume that \( V \) is distributed independent from \( u \),

\[
E(y) = 0 \quad \text{and} \quad E(V'V) = \psi.
\]

When incorporating prior information into the basic linear regression model, one obtains the mixed model in the form
\[
\begin{bmatrix}
y \\
r
\end{bmatrix} = 
\begin{bmatrix}
x \\
R
\end{bmatrix} B + 
\begin{bmatrix}
u \\
v
\end{bmatrix}
\]

where

\[
E \begin{bmatrix} u \\ v \end{bmatrix} = 0 \quad \text{and} \quad E \begin{bmatrix} u'v' \\ v \\ v \end{bmatrix} = 
\begin{bmatrix}
\sigma^2 I_n & 0 \\
0 & \psi
\end{bmatrix}
\]

(The off diagonal elements were zeros because of the assumption of independency of the distribution of \(v\) from the distribution of \(u\).)

Instead of OLS estimates \(\hat{\beta} = (x'x)^{-1}x'y\)

\[
\hat{b} = (x'R') \left( \begin{bmatrix}
\sigma^2 I_n & 0 \\
0 & \psi
\end{bmatrix} \right)^{-1} \begin{bmatrix} x' \\ R \end{bmatrix} \left( \begin{bmatrix}
\sigma^2 I_n & 0 \\
0 & \psi
\end{bmatrix} \right)^{-1} \begin{bmatrix} y \\ r \end{bmatrix}
\]

When \(\phi = \frac{1}{\sigma^2}\), \(\hat{b}\) becomes

\[
\hat{b} = (\phi x'x + R'\psi^{-1}R)^{-1} (\phi x'y + R'\psi^{-1}r)
\]

and

\[
\text{var - cov}(\hat{b}) = (\phi x'x + R'\psi^{-1}R)^{-1}
\]

Johnston (14) points out that the suggested Theil-Goldberger mixed model raises several questions.

(1) Is \(\hat{b}\) BLUE only in terms of the combined information, namely the prior information and the sample information (or in terms of the model \(V\) and \(x\) from the sample, and \(r\) and \(R\) from prior knowledge)?

(2) The Theil-Goldberger method was designed to cope with apparent conflict between sample and prior information and as a
result, \( \hat{b} \) will be a weighted average of the two sets of information, and it makes it difficult to separate just how much both sample and prior information each contribute to the estimate.

(3) The formula for estimating \( \hat{b} \) contains the term \( (\varphi - \frac{1}{\sigma^2}) \) and \( \sigma^2 \) is not usually known from the sample information.

The use of \( \hat{\sigma}^2 \) as an estimate for \( \sigma^2 \) or as recommended by Theil

\[
\hat{\sigma}^2 = \frac{(y'y - yx((x'x)^{-1})^{-1}x'y)}{n - k}
\]

will have only asymptotic distribution and has many limitations in generalization.

Sometimes the estimates are affected by dominancy of prior information and the sample information contribution is very limited. When this happens, the whole purpose of the regression analysis is undermined.

Prior Information in the Form of Inequality Restraints

G. G. Judge and T. Takayama (15) proposed a method to incorporate prior information in the form of inequality restraints that cannot be utilized with the Theil-Goldberger mixed model technique. This method can be summarized using their notation as follows

\[
y = x\beta + u, \quad E(u) = 0 \quad \text{and} \quad E(u'u) = \sigma^2 ln
\]

all as designed before
\hat{\beta} = (x'x)^{-1} x'y \quad \text{and} \quad V(\hat{\beta}) = \sigma^2(x'x)^{-1}.

When prior information is in the form of equality restraints, exact linear relationship is in the form \( r = R\beta \) where \( R \) is \((j \times k)\) constant known coefficient matrix, \( r \) is \((j \times 1)\) known vector. To find a vector \( \hat{b} \) which minimizes \( u'u = (y-x\beta)'(y-x\beta) \) and is subject to \( r - RB = 0 \) a solution by the lagrangian method yields

\[
\hat{b} = \beta + (x'x)^{-1} R' [R(x'x)^{-1} R']^{-1} (r - R\beta)
\]

\( \hat{b} \), the restricted estimate, is the best linear unbiased estimator (BLUE), and a linear function of \( y \) and \( r \).

When the prior information is not exact, then \( r = RB + V \) which is the case discussed by the Theil and Goldberger mixed model.

Assume that the following constraints prevail

\[
\begin{bmatrix}
I_1 \\
-I_1
\end{bmatrix}
\begin{bmatrix}
\beta_1
\end{bmatrix} \leq 
\begin{bmatrix}
r_u^1 \\
r_s^1
\end{bmatrix}
\quad \text{for} \quad 
\begin{bmatrix}
r^u_1 \\
r^s_1
\end{bmatrix} \geq 0 \quad \text{or} \quad 0 \leq r^3_1 \leq \beta_1 r^u_1
\quad (1)
\]

\[
\begin{bmatrix}
I_2 \\
-I_2
\end{bmatrix}
\begin{bmatrix}
\beta_2
\end{bmatrix} \leq 
\begin{bmatrix}
r_u^2 \\
r_s^2
\end{bmatrix}
\quad \text{for} \quad 
0 \leq r^s_2 \leq \beta_2 \leq r^u_2
\quad (2)
\]

\[
\begin{bmatrix}
I_3 \\
-I_3
\end{bmatrix}
\begin{bmatrix}
\beta_3
\end{bmatrix} \leq 
\begin{bmatrix}
r_u^3 \\
r_s^3
\end{bmatrix}
\quad \text{for} \quad 
0 \leq r^s_3 \leq \beta_3 \leq r^u_3
\quad (3)
\]

Where \( r^u_i \) and \( r^s_i \), \( i = 1, 2, 3 \) are known vectors of upper and lower bound constraints on the unknown coefficients in the \( i \)th set \( \beta_i \).
I for i = 1, 2, 3 is the identity matrix with a rank equal to the number of elements included in the parameter vector $\beta_i$. To find a $\hat{b}$ that minimize $u'u = (y - x\beta)'(y - x\beta)$ subject to the given constraints 1, 2, 3. The mathematical solution using Lagrangean multipliers can be seen in Judge and Takayama (15).

**Use of Ridge Regression**

A different way of incorporating prior information into the regression analysis is through the use of "ridge regression." The basic idea of ridge regression is that multicollinearity is moderated or alleviated by augmenting the main diagonal elements of the correlation matrix by small positive quantities.

Following Hoerl and Kennard's (10) notation, assume a linear model in a deviation form $y = x\beta + u$

where

$y$ denotes (nx1) vector of dependent variable,

$x$ is (nxp) matrix of explanatory variables. $\beta$ is (px1) vector of parameters, and $u$ is (nx1) vector of disturbances.

Assume $u$ has $E(u) = 0$

$$E(u'u) = \sigma^2 (In) = V$$

is positive definite.\(^7\)

\[^7\] V is a positive definite matrix if it is non-singular, has positive eigenvalues and a positive determinant.
Let \((x'x)\) denote the correlation matrix of the explanatory variables. \(\hat{\beta}^*\), the standardized ridge regression estimate of the coefficients is
\[
\hat{\beta}^* = (x'x + K\lambda)^{-1} x'y
\]
\[
\hat{\beta}^* = \beta + \gamma + (x'x + K\lambda)^{-1} x'u
\]
and variance-covariance of \(\hat{\beta}^*\) is
\[
\text{var - cov} (\hat{\beta}^*) = (x'x + K\lambda)^{-1} x'x (x'x + K\lambda)^{-1}
\]
The non-standardized \(\hat{\beta}^*\) estimate in terms of the corrected sums and cross-products is
\[
\hat{\beta}^* = (x'x + K\lambda)^{-1} x'y
\]
where \((x'x)\) is the \((pxp)\) matrix of mean corrected sum of squares, and cross-products, and
\[
\lambda = \text{diag} (\lambda_1, \lambda_2, \ldots, \lambda_p)
\]
the diagonal matrix of order \((p)\) consisting of the sums of squares.
\[
\text{var - cov} (\hat{\beta}^*)
\]
non-standardized is
\[
\text{var - cov} (\hat{\beta}^*) = \sigma^2 (x'x + K\lambda)^{-1} x'x (x'x + K\lambda)^{-1}
\]
Since ridge regression is a biased estimator, the expected bias of it is obtained from
\[
E(\hat{\beta}^*) = (x'x + K\lambda)^{-1} (x'x)\beta = (x'x + K\lambda)^{-1} [(x'x + K\lambda) - K\lambda] \beta
\]
\[
= (x'x + K\lambda)^{-1} (x'x + K\lambda)^{-1} \beta - (x'x + K\lambda)^{-1}(K\lambda) \beta
\]
\[
= \beta - K (x'x + K\lambda)^{-1} \lambda \beta
\]
where the bias of \( (\beta^*) \) is \(-K(x'x + K\lambda)^{-1}\lambda\). For more details, see Hoerl and Kennard (10) or Brown (1).
III. BIASED LINEAR ESTIMATIONS

In multiple linear regression models, ordinary least squares estimates (OLS) in the form \( \hat{\beta} = (x'x)^{-1} x'y \) are in the class of unbiased linear estimators (BLUE). But as discussed earlier, this class of estimators can have very undesirable and unacceptable characteristics. This is because of the large variance inflations of the estimates of the coefficients \( \beta \) that can occur because of multicollinearity.

In this section, a class of linear biased estimators will be briefly discussed.

Stein Estimators

The Shrunken Least Square Estimators

Assume a standard linear model in the form \( y = x\beta + u \) where as before \( y \) is \((nx1)\) vector of dependent variables, \( x \) is \((nxp)\) matrix of explanatory variables, \( \beta \) is \((px1)\) vector of parameters, and \( U \) is \((nx1)\) vector of disturbances

\[ u \sim N(0, \sigma^2 I) \] i.e., \( E(u) = 0, \ E(u'u) = \sigma^2 I \).

Let \((x'x)\) denote the correlation matrix, and \( \hat{\beta} = (x'x)^{-1} x'y \) the OLS estimates of the coefficients as before.

\[ \text{var - cov} (\hat{\beta}) = \sigma^2 (x'x)^{-1} \]
\[ \hat{\beta} \text{ has the distribution independent, normal, with a mean } \beta \]
and \[ \sigma^2 (x'x)^{-1} \text{ variance, } \hat{\beta} \sim N \left( \beta, \sigma^2 (x'x)^{-1} \right) \]

Assume the following transformation for simplicity: \( Q \) is an Orthogonal symmetric positive definite matrix, such that

\[ Q'Q = QQ' = I \]
\[ Q'(x'x)Q = \Lambda = \text{diag} \left( \lambda_1, \lambda_2, \ldots, \lambda_p \right) \]

where \( (x'x) = \text{correlation matrix} \) and \( \lambda_i \) is the eigenvalue of \( (x'x) \).

Let \( Z = Qx \) and \( \alpha = Q\beta \). Then the linear model is transformed to \( y = Z\alpha + e \). As before

\[ e \sim N(0, \sigma^2 I_n) \]
\[ \alpha \sim N(\alpha, \sigma^2 / \lambda_1) \]
\[ \hat{\beta}^* = (x'x + KT)^{-1} x'y \]

\( T \) is a positive definite symmetric matrix and \( K \) is a non-negative scalar.

Stein estimators are obtained by choosing \( T = (x'x) \) and \( \hat{\alpha}_s^* \)

Stein estimates will be equal to \( \hat{\alpha}_s^* = f \alpha_i \), where \( f = (1/1+K) \) that

gives all the coefficients an equal weight, which is a function of \( (K) \).

Stein M \(^8/\), the most promising in the Stein class, in which

\[ \alpha_i^* \sim N \left( f \alpha_i, \ f^2 \sigma^2 / \lambda_i \right) \]

\( K \) is chosen so that \( \sum \lambda_i \hat{\alpha}_i^2 / (w^2 + \sigma^2) \) is equal to its marginal

\(^8/\) Stein M was first introduced and evaluated by Dempster, Scharzoff and Wermuth in 1977.
expectation \( p \) where \( \hat{\sigma}^2 \) is substituted for \( \sigma^2 \) and \( w^2 = \sigma^2 / K \).

Because of the fact that \( f = (1/1+K) \) in Stein estimates is not a function of the sample in formation \( (x'x) \), Stein estimators have very little to recommend them for alleviating the multicollinearity problem.

**Ridge Regression**

Based on the choice of \( K \) in ridge regression, a class of ridge estimators is defined, i.e., each choice of \( (K) \) determines a ridge estimator.

Three of this class of biased estimators will be discussed in this section.

**Hoerl-Baldwin and Kennerd Ridge Regression**

Hoerl-Baldwin and Kennerd (11) proposed a technique for choosing \( K \). In this paper, \( K \) will be referred to as \( K_A \).

As before, assume a general linear model \( y = x\beta + u \) where \( x \) is \((nxp)\) matrix of variables, \( y \) is \((nx1)\) vector of observations, \( \beta \) is \((px1)\) vector of parameters, and \( u \) is \((nx1)\) vector of disturbances. \( u \sim N(0, \sigma^2 I) \) and \( (x'x) \) is a correlation matrix.

Assume the same transformation applies

\[
\hat{\beta}^* = (x'x + Q'KQ)^{-1} x'y
\]

where \( K \) is a non-negative diagonal matrix and \( (x'x) = Q' \lambda Q \).

\( \lambda \) is the diagonal matrix of eigenvalues.
If $\alpha = Q_{\beta}$ minimum MSE^(2)^\(^2/\) (Mean Squared Error) is obtained when $K_1 = \sigma^2 / \alpha^2_1$ and the harmonic mean of these individual is

\[
\left( \frac{1}{K_A} \right) = \left( \frac{1}{\sigma^2} \right) \Sigma \alpha^2 = \frac{\alpha^2}{\rho^\sigma^2}
\]

\[
= \beta^2 / \rho^\sigma^2
\]

\[
K_A = \frac{\rho^\sigma^2}{\beta^2}
\]

and since $\sigma^2$ and $\beta$ are unknown, then $K_a$ can be estimated as

\[
K_A = \frac{p^\sigma^2}{\hat{\beta}^2}
\]

where $\hat{\beta}$ is OLS estimate for $\beta$, and $\hat{\sigma}^2$ is also an OLS estimate for $\sigma^2$.

$K_A$ defines HBK ridge regression.

Lawless and Wang Ridge Estimators

This estimator is the same as HBK $K_A$ estimators, except that the coefficients are weighted according to the eigenvalues of $(x'x)$. $K_B$ is chosen in such a way that

\[
K_B = \frac{\rho^\sigma^2}{\rho^p} \Sigma \lambda_i \alpha^2_i
\]

\[
i=1
\]

where $p$ denotes the number of explanatory variables, $\lambda_i$ is the $i^{th}$ eigenvalue of $(x'x)$, $\alpha_i$ is the OLS estimates of the coefficients and $\hat{\sigma}^2$ is the OLS estimate of $\sigma^2$.

\[
MSE(\hat{\beta}) = V(\hat{\beta}) + [E(\hat{\beta} - \beta)]^2
\]
In a study by Lawless and Wang (16) this choice of $K_B$ proved to be more effective and produced better results than under the use of $K_A$.

RIDGM

RIDGM developed by Dempster, Shatzoff and Wermuth (5) is motivated by the Bayesian interpretation of ridge. The observable least square estimate $\hat{\alpha}_i$ is marginally independently distributed $N(0, \sigma^2/\lambda_i)$.

$K_M$ is chosen in such a way that the prior expectation of

$$\sum \hat{\alpha}_i^2/w^2 + \sigma^2/\lambda_i$$

is $p$ where $\hat{\alpha}_i$ is the OLS estimated of the coefficients, $\sigma^2$ is substituted for $\sigma$, $\lambda_i$ is the $i^{th}$ eigenvalue of $(x'x)$, and $w^2$ is $\sigma^2/K_M$ and $p$ is the number of parameters. This specific choice of $K$ yields a ridge regression referred to as RIDGM.
IV. APPLICATION OF SOME RIDGE REGRESSION MODELS FOR THE ESTIMATION OF ECONOMIC RELATIONSHIPS

Nature of Economic Data

Unfortunately, economists seldom have good quality data of primary source. Instead, one often turns to some secondary sources, such as the U.S. Census Agricultural and Business Survey reports. These provide a somewhat poor indication of many important inputs in our economic models.

Data are always available on an aggregate basis, such as counties and states, and one has to use these data in spite of the fact that aggregation results in a loss of information and a higher degree of multicollinearity.

To solve the data problems, it is better to return to the original data whenever possible. When such a return is not possible, different techniques to cope with the problem are tried.

In this section, several economic problems are analyzed using ridge models. These economic models are chosen mainly because:

1. They suffer from multicollinearity, i.e., they lack independence in one or more of their explanatory variables.

2. These models have some symptoms of nonsense regression or harmful multicollinearity, i.e., the coefficients may take wrong signs or have low t values.
These models cannot be handled very well by ordinary least square regression, and one or more of their important explanatory variables is not statistically different from zero when OLS is used.

The analysis results of ridge regression will be viewed in terms of their performances on a theoretical basis and expectation, and how they compare with ordinary least square estimates.

Ridge Regression in Estimating the Marginal Value Productivity of Irrigation Water

Background

Using U.S. Census Agricultural data, Ruttan pioneered the estimation of total value product functions to derive an economic value for the water used in irrigating agriculture. The original formulation of the production function model included six important explanatory variables (1).

In functional relationship, the model was

\[ y = g(x_1, x_2, x_3, x_4, x_5, x_6) \]

where

- \( y \) denotes farm product sold;
- \( x_1 \) is number of family and hired labor;
- \( x_2 \) is number of tractors on the farm;
- \( x_3 \) is value of livestock investment in the farm;
- \( x_4 \) is acres of irrigated cropland;
Table 1. Estimated values for regression of county values of all farm products sold as a Cobb-Douglas function of inputs, OLS estimates for 25 Central Pacific counties, 1954.

<table>
<thead>
<tr>
<th></th>
<th>$X_1$</th>
<th>$X_2$</th>
<th>$X_3$</th>
<th>$X_4$</th>
<th>$X_5$</th>
<th>$X_6$</th>
<th>$y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Means</td>
<td>4.03550944</td>
<td>3.589889</td>
<td>7.0501228</td>
<td>5.249523</td>
<td>3.168056</td>
<td>7.0311667</td>
<td>7.7910822</td>
</tr>
<tr>
<td>Betas (OLS)</td>
<td>0.269805</td>
<td>-0.0628507</td>
<td>0.0227421</td>
<td>0.420649</td>
<td>0.155389</td>
<td>0.521785</td>
<td></td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>0.215133</td>
<td>0.279637</td>
<td>0.145926</td>
<td>0.145286</td>
<td>0.0900693</td>
<td>0.151354</td>
<td></td>
</tr>
<tr>
<td>t values</td>
<td>1.25413</td>
<td>-2.24758</td>
<td>1.155847</td>
<td>2.87553</td>
<td>1.72522</td>
<td>3.44745</td>
<td></td>
</tr>
</tbody>
</table>

$X_1$ = number of family and hired workers  
$X_2$ = number of tractors  
$X_3$ = value of livestock investment  
$X_4$ = acres of irrigated cropland  
$X_5$ = acres of non-irrigated cropland  
$X_6$ = current operating expenses  
$y$ = farm products sold  
$R^2 = 0.925$
\( x_5 \) is acres of non-irrigated cropland, and 

\( x_6 \) is current operating expenditure.

The Cobb-Douglas production function was fitted to the model. Means, OLS estimates, standard deviation and \( t \) values are presented in Table 1.

**Multicollinearity in the Model**

The correlation matrix given in Table 2 shows that

\[
\begin{align*}
r_{12} &= 0.9299 \\
r_{13} &= 0.7624 \\
r_{14} &= 0.5968 \\
r_{16} &= 0.8188
\end{align*}
\]

which indicates a degree of collinearity between \( x_1 \) and the remaining variable set.

Also from Table 2, \( x_2 \) is highly collinear with some variable sets in the model.

\[
\begin{align*}
r_{23} &= 0.8048 \\
r_{24} &= 0.7685 \\
r_{26} &= 0.80425
\end{align*}
\]

The main diagonal elements of the inverted correlation matrix were:
Table 2. Simple correlation coefficient for regression of county values of all farm products sold as a Cobb-Douglas function of inputs.

<table>
<thead>
<tr>
<th></th>
<th>$X_1$</th>
<th>$X_2$</th>
<th>$X_3$</th>
<th>$X_4$</th>
<th>$X_5$</th>
<th>$X_6$</th>
<th>$y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_1$</td>
<td>1.00000000</td>
<td>0.929918569</td>
<td>0.762442321</td>
<td>0.59685312</td>
<td>-0.293808879</td>
<td>0.811853556</td>
<td>0.857723175</td>
</tr>
<tr>
<td>$X_2$</td>
<td>1.00000000</td>
<td>0.804854828</td>
<td>0.768523668</td>
<td>-0.343767007</td>
<td>0.804250815</td>
<td>0.895857998</td>
<td></td>
</tr>
<tr>
<td>$X_3$</td>
<td>1.00000000</td>
<td>0.652256285</td>
<td>-0.394812482</td>
<td>0.849321658</td>
<td>0.834051825</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$X_4$</td>
<td>1.00000000</td>
<td>0.598540061</td>
<td>-0.507591052</td>
<td>-0.251882809</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$X_5$</td>
<td>1.00000000</td>
<td>-0.170285996</td>
<td>-0.251882809</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$X_6$</td>
<td>1.00000000</td>
<td>0.896652600</td>
<td>1.00000000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$y$</td>
<td>1.00000000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Variable | Main diagonal or VIF
---|---
$x_1$ | 11.2
$x_2$ | 18.9
$x_3$ | 05.2
$x_4$ | 05.2
$x_5$ | 02.0
$x_6$ | 05.5

These diagonal elements show that $x_2$ is the most likely to be affected by multicollinearity and $x_1$ is in the second place. $x_1$ and $x_2$ were highly correlated with the other variables.

Since

$$r_{213456} = 1 - (1 \times 18.9) = 0.947$$

and

$$r_{123456} = 1 - (1 \times 11.2) = 0.9107,$$

then multicollinearity, brought about by the practical problem of separating the effects of $x_1$ (labor) from ($x_2$) tractors on productivity, leads to some serious problems in the model when ordinary least square is fitted. Some of the problems are:

1. Symptoms of harmful multicollinearity, since $\hat{\beta}_2$ took an illogical negative sign.

2. Four out of six explanatory variables in the model, namely $x_1, x_2, x_3$ and $x_5$, were statistically insignificant at the 5% confidence level, due to the high $V(\hat{\beta}_1)$. 
(3) Only $x_4$ and $x_6$ were statistically significant at the 5% probability level.

Due to these serious problems caused by multicollinearity, many econometricians advocate the idea of deletion of the variables that were highly affected by the problem, and only a subset of these variables was used in the formulation of the final model.

To avoid the possibly serious bias introduced by the specification error which results from the deletion of important variables, the suggested ridge regression (with $K_A$, $K_B$, and $K_M$) was fitted to the basic data using the same correlation matrix.

The coefficients, variances and their estimated MSE ($\hat{\beta}_1$) for OLS, $K_A$, $K_B$ and $K_M$, are given in Table 3. Very important results are noticed in the table and these are:

1. $\hat{\beta}_2$ has changed from an unexpected negative sign ($-0.06285$) to an expected positive sign.
   
   $\hat{\beta}_2^*$ = 0.1056 when $K_A$ is used
   $\hat{\beta}_1^*$ = 0.1275 when $K_M$ is used; and,
   $\hat{\beta}_2^*$ = 0.0699 using $K_B$ ridge regression.

2. All the coefficients in the model were possibly statistically significant at 5% (if the problem of bias is temporarily neglected) due to the fact that more precision was obtained and the variances of $V(\hat{\beta}_1)$ were reduced (see Table 4).
Table 3. Estimates of the betas, variances and estimates MSE, OLS vs. K, KM and KA for regression of county values of all farm products sold as a Cobb-Douglas function of inputs.

<table>
<thead>
<tr>
<th></th>
<th>K = 0</th>
<th>Lawless-Wang K</th>
<th>RIDGM</th>
<th>HBK</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{\beta}_1 )</td>
<td>0.2698</td>
<td>0.2015</td>
<td>0.1821</td>
<td>0.1882</td>
</tr>
<tr>
<td>V(( \hat{\beta}_1 ))</td>
<td>0.0462824</td>
<td>0.0167051</td>
<td>0.00874386</td>
<td>0.0114868</td>
</tr>
<tr>
<td>V(( \hat{\beta}_1 )) + E(( \hat{\beta}_1 ) - ( \hat{\beta}_1 ))^2</td>
<td>0.0213712</td>
<td>0.0164372</td>
<td>0.0181405</td>
<td></td>
</tr>
<tr>
<td>( \hat{\beta}_2 )</td>
<td>-0.06285</td>
<td>0.0699</td>
<td>0.1275</td>
<td>0.1056</td>
</tr>
<tr>
<td>V(( \hat{\beta}_2 ))</td>
<td>0.0891969</td>
<td>0.0226519</td>
<td>0.00939931</td>
<td>0.0137192</td>
</tr>
<tr>
<td>V(( \hat{\beta}_2 )) + E(( \hat{\beta}_2 ) - ( \hat{\beta}_2 ))^2</td>
<td>0.040276</td>
<td>0.0456276</td>
<td>0.0420975</td>
<td></td>
</tr>
<tr>
<td>( \hat{\beta}_3 )</td>
<td>0.02274</td>
<td>0.0613</td>
<td>0.0935</td>
<td>0.07952</td>
</tr>
<tr>
<td>V(( \hat{\beta}_3 ))</td>
<td>0.0212945</td>
<td>0.0141309</td>
<td>0.00939659</td>
<td>0.0113570</td>
</tr>
<tr>
<td>V(( \hat{\beta}_3 )) + E(( \hat{\beta}_3 ) - ( \hat{\beta}_3 ))^2</td>
<td>0.0156170</td>
<td>0.0144069</td>
<td>0.0145812</td>
<td></td>
</tr>
<tr>
<td>( \hat{\beta}_4 )</td>
<td>0.42065</td>
<td>0.3435</td>
<td>0.2969</td>
<td>0.3165</td>
</tr>
<tr>
<td>V(( \hat{\beta}_4 ))</td>
<td>0.0213995</td>
<td>0.019185</td>
<td>0.007310094</td>
<td>0.00867318</td>
</tr>
<tr>
<td>V(( \hat{\beta}_4 )) + E(( \hat{\beta}_4 ) - ( \hat{\beta}_4 ))^2</td>
<td>0.168770</td>
<td>0.0226222</td>
<td>0.0195232</td>
<td></td>
</tr>
<tr>
<td>( \hat{\beta}_5 )</td>
<td>0.15539</td>
<td>0.1353</td>
<td>0.1211</td>
<td>0.127550</td>
</tr>
<tr>
<td>V(( \hat{\beta}_5 ))</td>
<td>0.00811247</td>
<td>0.00618171</td>
<td>0.00304492</td>
<td>0.00553259</td>
</tr>
<tr>
<td>V(( \hat{\beta}_5 )) + E(( \hat{\beta}_5 ) - ( \hat{\beta}_5 ))^2</td>
<td>0.00658359</td>
<td>0.00621976</td>
<td>0.00630762</td>
<td></td>
</tr>
<tr>
<td>( \hat{\beta}_6 )</td>
<td>0.52179</td>
<td>0.4597</td>
<td>0.4082</td>
<td>0.4309</td>
</tr>
<tr>
<td>V(( \hat{\beta}_6 ))</td>
<td>0.0229081</td>
<td>0.0146759</td>
<td>0.00938154</td>
<td>0.0115460</td>
</tr>
<tr>
<td>V(( \hat{\beta}_6 )) + E(( \hat{\beta}_6 ) - ( \hat{\beta}_6 ))^2</td>
<td>0.0185291</td>
<td>0.0222841</td>
<td>0.0198085</td>
<td></td>
</tr>
</tbody>
</table>

* Ordinary least square estimates \( \hat{\beta} \) were used to estimate MSE(\( \hat{\beta}_1 \) *).
Table 4. Variances and estimated MSE of coefficients OLS vs. $K_A$, $K_B$ and $K_M$ (OLS = 100.00) for regression of county values of all farm products sold as a Cobb-Douglas function of inputs for 25 Central Pacific counties, 1954.

<table>
<thead>
<tr>
<th></th>
<th>OLS K = 0</th>
<th>$K_\beta$</th>
<th>$K_M$</th>
<th>$K_A$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V(\hat{\beta}_1)$</td>
<td>100.00</td>
<td>36.09%</td>
<td>18.89%</td>
<td>24.82</td>
</tr>
<tr>
<td>$V(\hat{\beta}_1) + E(\hat{\beta}_1 - \beta_1)^2$</td>
<td>100.00</td>
<td>46.17</td>
<td>35.5</td>
<td>39.2</td>
</tr>
<tr>
<td>$V(\hat{\beta}_2)$</td>
<td>100.00</td>
<td>28.96</td>
<td>12.02</td>
<td>17.54</td>
</tr>
<tr>
<td>$V(\hat{\beta}_2) + E(\hat{\beta}_2 - \beta_2)^2$</td>
<td>100.00</td>
<td>51.5</td>
<td>58.34</td>
<td>53.84</td>
</tr>
<tr>
<td>$V(\hat{\beta}_3)$</td>
<td>100.00</td>
<td>66.35</td>
<td>44.13</td>
<td>53.33</td>
</tr>
<tr>
<td>$V(\hat{\beta}_3) + E(\hat{\beta}_3 - \beta_3)^2$</td>
<td>100.00</td>
<td>73.34</td>
<td>67.66</td>
<td>68.47</td>
</tr>
<tr>
<td>$V(\hat{\beta}_4)$</td>
<td>100.00</td>
<td>51.02</td>
<td>34.16</td>
<td>40.53</td>
</tr>
<tr>
<td>$V(\hat{\beta}_4) + E(\hat{\beta}_4 - \beta_4)^2$</td>
<td>100.00</td>
<td>78.87</td>
<td>105.7</td>
<td>91.23</td>
</tr>
<tr>
<td>$V(\hat{\beta}_5)$</td>
<td>100.00</td>
<td>76.20</td>
<td>62.19</td>
<td>68.19</td>
</tr>
<tr>
<td>$V(\hat{\beta}_5) + E(\hat{\beta}_5 - \beta_5)^2$</td>
<td>100.00</td>
<td>81.15</td>
<td>76.67</td>
<td>77.75</td>
</tr>
<tr>
<td>$V(\hat{\beta}_6)$</td>
<td>100.00</td>
<td>64.06</td>
<td>40.95</td>
<td>50.40</td>
</tr>
<tr>
<td>$V(\hat{\beta}_6) + E(\hat{\beta}_6 - \beta_6)^2$</td>
<td>100.00</td>
<td>80.88</td>
<td>97.27</td>
<td>86.47</td>
</tr>
<tr>
<td>Over all</td>
<td>100.00</td>
<td>60.17</td>
<td>64.38</td>
<td>60.77</td>
</tr>
</tbody>
</table>

* $V(\beta^*) + [E(\hat{\beta} - \beta)]^2$ was used to estimate MSE ($\beta^*$) because the true $\beta$ values are unknown.
was reduced 50-80% from the original under OLS. KM yielded the highest reduction in variance, \( V(\hat{\beta}_1) \) reduced 70-88%. Again, KM performs the best if possible bias is ignored. In \( V(\hat{\beta}_4) \) there was 50-60% reduction variance, and \( V(\hat{\beta}_6), V(\hat{\beta}_5) \) was down by 30-60%.

Among the three ridge regressions used in this analysis, all performed well. The overall performance of the three ridge regression models were obviously superior to that of ordinary least square procedure in terms of the coefficient signs and the precision of the estimates. This type of problem seems to be better tackled by ridge regression rather than OLS.

**A Production Function Analysis of Water Resource Productivity in Pacific Northwest Agriculture**

**Background**

The basic data for this analysis was taken from M. H. Holloway (13) (Appendix Table IV). Agriculture has been a major consumer of water in the Pacific Northwest, and the basic objective of this study is to determine the contribution of agricultural water resource development to recent agricultural production.

Cobb-Douglas production function was selected and ordinary least square procedure was primarily chosen to determine the initial estimates.
Table 5. Ordinary least square estimates, standard errors, means and t-values for analysis of water resource productivity in Pacific Northwest (Area C) 1964.

<table>
<thead>
<tr>
<th></th>
<th>(X₁)</th>
<th>(X₂)</th>
<th>(X₃)</th>
<th>(X₄)</th>
<th>(X₅)</th>
<th>(X₆)</th>
<th>(X₇)</th>
<th>(X₈)</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLS (β₁)</td>
<td>-5.0295</td>
<td>2.7431</td>
<td>1.2982</td>
<td>8.4439</td>
<td>-2.1119</td>
<td>9.1729</td>
<td>16.2737</td>
<td>-1.9993</td>
<td></td>
</tr>
<tr>
<td>S E(β₁)</td>
<td>0.2369</td>
<td>0.1422</td>
<td>0.3119</td>
<td>0.1341</td>
<td>0.0781</td>
<td>0.1174</td>
<td>0.08408</td>
<td>0.1016</td>
<td></td>
</tr>
<tr>
<td>means Xᵢ</td>
<td>1241.10</td>
<td>5236.29</td>
<td>3889.34</td>
<td>466.678</td>
<td>180.191</td>
<td>225.613</td>
<td>10.7535</td>
<td>84.468</td>
<td></td>
</tr>
<tr>
<td>T values</td>
<td>-1.21953</td>
<td>5.22212</td>
<td>0.7946</td>
<td>2.2225</td>
<td>-3.68790</td>
<td>2.463536</td>
<td>2.29539</td>
<td>-3.92828</td>
<td></td>
</tr>
</tbody>
</table>


Area of Analysis

Only area 'C' which consists of 20 counties producing mostly field crops will be analyzed.

The Model

Cobb-Douglas production function (linear form) was fitted.

\[ y = F(x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8) \]

where

- \( y \) denotes values of farm products sold and value of home consumption.
- \( x_1 \) is hired and family labor;
- \( x_2 \) is value of current operating expenses;
- \( x_3 \) is flow of capital on farm;
- \( x_4 \) is cropland quantity adjusted for quality;
- \( x_5 \) is animal unit months;
- \( x_6 \) is irrigation water application;
- \( x_7 \) is service flow of form investment in drainage;
- \( x_8 \) is service flow of farm investment in water conservation practice.

Ordinary least square estimates, standard deviations, \( t \) values, and means of the variables are given in Table 5.
Table 6. Water resource productivity in Pacific Northwest agriculture. Simple correlation between variables.

<table>
<thead>
<tr>
<th></th>
<th>$X_1$</th>
<th>$X_2$</th>
<th>$X_3$</th>
<th>$X_4$</th>
<th>$X_5$</th>
<th>$X_6$</th>
<th>$X_7$</th>
<th>$X_8$</th>
<th>$Y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_1$</td>
<td>1.0000000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$X_2$</td>
<td>.9357189</td>
<td>1.0000000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$X_3$</td>
<td>.9550682</td>
<td>.9138426</td>
<td>1.0000000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$X_4$</td>
<td>.4227123</td>
<td>.5606500</td>
<td>.5220780</td>
<td>1.0000000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$X_5$</td>
<td>.5271189</td>
<td>.4865410</td>
<td>.4759668</td>
<td>.0192848</td>
<td>1.0000000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$X_6$</td>
<td>.4992776</td>
<td>.3576612</td>
<td>.4362811</td>
<td>-0.2761547</td>
<td>.2773230</td>
<td>1.0000000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$X_7$</td>
<td>.3263360</td>
<td>.4369850</td>
<td>.4825480</td>
<td>.7575409</td>
<td>.1787583</td>
<td>.3354978</td>
<td>1.0000000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$X_8$</td>
<td>.3316825</td>
<td>.3137947</td>
<td>.4439894</td>
<td>.1110539</td>
<td>-0.0948184</td>
<td>.1058459</td>
<td>.2659048</td>
<td>1.0000000</td>
<td></td>
</tr>
<tr>
<td>$Y$</td>
<td>.8908467</td>
<td>.9510433</td>
<td>.9101004</td>
<td>.6511569</td>
<td>.3913535</td>
<td>.4263892</td>
<td>.4818734</td>
<td>.2175557</td>
<td>1.0000000</td>
</tr>
</tbody>
</table>
**Multicollinearity**

With reference to the sample correlation matrix, (Table 6) and to the diagonal elements of the inverted correlation matrix (VIF), Table 7, the indication is that $x_1$, $x_2$, and $x_3$ are the most affected by multicollinearity since $x_3$ has the highest value of 59.487, $x_1$ is 32.584 and $x_2$ has a value of 11.732. Also, one notices that $r_{12} = 0.9357$ and $r_{13} = 0.955$, which points to the high degree of dependency between these three variables.

Multicollinearity is brought about by the dependency between some of these variables in the variable sets and it is due to the loss of information because of the aggregation in measuring the variables. Multicollinearity here is unavoidable.

**Consequences of Multicollinearity in the Model**

1. Ordinary least square procedure yields estimates which are unstable because the variances of the estimates are unreasonably high.

2. $\hat{\beta}_1$, $\hat{\beta}_5$, and $\hat{\beta}_8$ took illogical, negative signs. These signs were not expected theoretically, and were caused only by the presence of multicollinearity. This is especially true of $\hat{\beta}_1$ and $\hat{\beta}_5$, because $x_1$ and $x_5$ are highly collinear with other variables in the variable set.

3. Most of the variables in the model were statistically
Table 7. Variance inflation factors in the diagonal elements of the inverted correlation matrix (XX) Pacific Northwest water resource productivity data (Area C), 1964.

<table>
<thead>
<tr>
<th>Explanatory Variables</th>
<th>Main diagonal elements of inverted correlation matrix (VIF)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(X_1)$</td>
<td>32.584</td>
</tr>
<tr>
<td>$(X_2)$</td>
<td>11.732</td>
</tr>
<tr>
<td>$(X_3)$</td>
<td>56.487</td>
</tr>
<tr>
<td>$(X_4)$</td>
<td>10.436</td>
</tr>
<tr>
<td>$(X_5)$</td>
<td>03.538</td>
</tr>
<tr>
<td>$(X_6)$</td>
<td>08.006</td>
</tr>
<tr>
<td>$(X_7)$</td>
<td>04.105</td>
</tr>
<tr>
<td>$(X_8)$</td>
<td>05.997</td>
</tr>
<tr>
<td></td>
<td>$\hat{\beta}_1$</td>
</tr>
<tr>
<td>-------</td>
<td>-----------------</td>
</tr>
<tr>
<td>S.E.</td>
<td>(0.2369)</td>
</tr>
<tr>
<td>PM-1</td>
<td>0.917</td>
</tr>
<tr>
<td>S.E.</td>
<td>(0.3238)</td>
</tr>
<tr>
<td>rldgm</td>
<td>3.113</td>
</tr>
<tr>
<td>S.E.</td>
<td>(0.0168)</td>
</tr>
<tr>
<td>$K_B$</td>
<td>2.9638</td>
</tr>
<tr>
<td>S.E.</td>
<td>(0.02334)</td>
</tr>
<tr>
<td>$K_A$</td>
<td>0.904</td>
</tr>
<tr>
<td>S.E.</td>
<td>(0.0661)</td>
</tr>
</tbody>
</table>
insignificant and not different from zero at the 5% level (Table 8). This results from the fact that the $V(\hat{\beta}_1)$ are high and resulting t values are small.

Prior Information Model

Due to the problem of multicollinearity, Holloway (13) fitted another model based on incorporating prior information into the regression analysis. The Theil-Goldberger approach (14) was chosen among several alternatives because the model allows the specification of the prior parameter estimates with variances (13).

In this analysis, a comparison between ordinary least square, the Theil-Goldberger prior information model (discussed earlier) and ridge regression, will be made.

The Use of Ridge Regression

Although ridge regression can be thought of as another way of incorporating prior information into the regression analysis, it has some advantages over both OLS and the Theil-Goldberger prior information mixed model, as shown in Table 8.

Using the basic data given by Holloway (13), ridge regression in terms of $K_B$ and $K_M$ were fitted, and the results are given in Table 8.
Comparison Between Ordinary Least Square, Theil-Goldberger Prior Information Model and Ridge Regression Models

With reference to Table 8, it was noted that

1. \( \hat{\beta}_1 \) estimate changed from a wrong negative sign to a positive sign for the Theil-Goldberger prior information (PM-1)\(^1\) model. All ridge procedures (\( K_A \), \( K_B \) and \( K_M \)) seem to also correct for this problem, although the magnitude of the coefficient in \( K_A \) and Theil-Goldberger PM-1 were close (0.917 for PM-1 and 0.904 for \( K_A \)).

2. \( \hat{\beta}_5 \) also changed from a wrong negative sign under OLS to a positive sign in PM-1 and \( K_M \). For \( \hat{\beta}_8 \), the sign did not change for any of the ridge regression models \( K_A \) and \( K_B \).

Overall results indicate that the use of ridge \( m \) worked as well as the use of the Theil-Goldberger mixed model in correcting the wrong sign caused by multicollinearity, except for \( \beta_8 \).

\( \hat{\beta}_8 \) changed into positive sign in PM-1, as a result of the prior information dominancy over the sample information and in this case the sample contribution to the estimate was minimum.

2. The variances of the coefficients \( V(\hat{\beta}_1) \) were reduced under ridge regression with ridge performing the best, while PM-1 \( V(\hat{\beta}_1) \) was higher than that obtained by OLS. For example:

\[ SE(\hat{\beta}_1) = 0.2369 \text{ in OLS} = 0.628 \text{ in PM-1 and } 0.0168 \text{ in ridge m.} \]

\(^1\) Theil-Goldberger model will be referred to as PM-1 throughout this section.
the same relativity applies to all the coefficients in Table 8.

(3) All the coefficients were significant under ridge \( m \) at 5% confidence level, if one ignores the bias of the ridge \( m \) estimates.

This problem suffers from a severe case of multicollinearity, worsened by some degree of aggregation in the data. Ridge \( m \) seems to perform as well as the prior information model utilized by Holloway (13) to account for multicollinearity, and in fact it has very good merits as an effective tool in this type of problem.

Prior information about the coefficients and their variances is not always available in a practical situation. If they are available almost always there is a bias in utilizing this information because of the different circumstances under which it is obtained. Thus, the assumption of an unbiased prior estimate of one or more \( \beta \) values would often be violated in using the Theil-Goldberger mixed model.

The use of prior information as in this case was associated with the dominance of this prior information over the sample information, which raises questions about the results. Use of the Theil-Goldberger model under these conditions tends to reflect only the prior information effect, and the sample information contribution tends to be obscured by the dominancy of the possibly biased prior information.
Salmon Steelhead Fishing Demand Relationship

Basic data for this analysis was given by Brown, Singh and Castle (3) and by Brown (2).

A demand relationship was proposed in such a way that it differed from past studies. The difference was in its separation of two types of cost involved: a direct cost measured in a monetary form, and another type of cost, time cost, measured by the distance of travel in miles.

The demand functional relationships were

\[ \text{DYS} = F(\text{INC}, \text{MLS}, \text{CST}, j) \]

\( \text{DYS} \) denotes Salmon steelhead (S-S) days of fishing taken per unit of population of subzone \( j \).

\( \text{INC} \) is average family income of subzone \( j \)

\( \text{MLS} \) is average miles per (S-S) trip of subzone \( j \)

and \( \text{CST} \) is average (S-S) variable cost per day of subzone \( j \).

The demand function fitted by Ordinary Least Square (OLS) yields the following estimate:

\[ \ln \text{DYS} = 0.7054 \text{INC}_j + 0.2948 \text{MLS}_j - 1.1691 \text{CST}_j \]

(Standard deviations of the coefficients are reported in parentheses below the corresponding regression coefficients.)

\[ R^2 = 0.6534 \]
Table 9. Simple correlation coefficient matrix for salmon steelhead fishing demand relationship.

<table>
<thead>
<tr>
<th></th>
<th>INC</th>
<th>MLS</th>
<th>CST</th>
<th>DYS</th>
</tr>
</thead>
<tbody>
<tr>
<td>INC</td>
<td>1.000000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MLS</td>
<td>0.282220</td>
<td>1.000000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CST</td>
<td>0.493300</td>
<td>0.872860</td>
<td>1.000000</td>
<td></td>
</tr>
<tr>
<td>DYS</td>
<td>0.211790</td>
<td>-0.526645</td>
<td>-0.563864</td>
<td>1.000000</td>
</tr>
</tbody>
</table>

DYS = Salmon days of fishing
INC = Average family income
MLS = Average miles per S-S trip
CST = Average variable cost S-S per day
Multicollinearity

Multicollinearity brought about by the difficulty in separating the effect of the monetary cost of travel versus the time cost, is discussed by Brown (3). The use of zone average and the loss of information through the process of aggregation are some of the reasons to suspect multicollinearity.

Test for Multicollinearity

Examining the simple correlation matrix shown in Table 9, it is seen that the highest simple correlation is between the variables, MLS and CST = $r_{23} = 0.87286$.

A good indication of the degree of multicollinearity is given by the main diagonal elements of the inverted correlation matrix ($x'^*x$). These were

<table>
<thead>
<tr>
<th>Explanatory Variable</th>
<th>Variance Inflation Factor (VIF)</th>
</tr>
</thead>
<tbody>
<tr>
<td>INC,</td>
<td>1.5057</td>
</tr>
<tr>
<td>MLS</td>
<td>4.7844</td>
</tr>
<tr>
<td>CST</td>
<td>5.8197</td>
</tr>
</tbody>
</table>

The highest variance inflation factor's (VIF) corresponded to the two variables most affected by multicollinearity, namely MLS and CST.
As a consequence of multicollinearity, the following took place.

(1) Ordinary least square estimates ($\hat{\beta}_2$) the coefficient of MLS took an insignificant, wrong and theoretically unexpected positive sign.

(2) OLS coefficients, except for INC and CST, were statistically insignificant due to the relatively high variance associated with the coefficients.

Improving the Estimation of the Demand Relationship

To improve the estimation of the demand relationships, the three types of ridge regression suggested to cope with the problem of multicollinearity were applied to the data. The results are given in Tables 10 and 11. With reference to these tables, the following results are evident.

(1) Hoerl-Baldwin and Kennard Basic $K_A$ Performance

$\hat{\beta}_1$ changed from a large positive number to a relatively smaller positive number.

$\beta_2$ is still taking an illogical positive sign although some improvement in the $\hat{\text{V}}(\hat{\beta}_2)$ is noticed.

(2) Lawless and Wang $K_B$ Performances

(i) $\hat{\beta}_2$ changed from illogical positive to the theoretically expected negative sign. $\hat{\beta}_2$ changed from 0.29468 to -.038079 so the sign has been corrected.
Table 10. Salmon steelhead demand relationship coefficients, variances and estimated MSE, OLS vs. ridge regression.

<table>
<thead>
<tr>
<th></th>
<th>OLS</th>
<th>HGK</th>
<th>LW</th>
<th>RIDGM</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_1^*$</td>
<td>0.705293</td>
<td>0.591099</td>
<td>0.533813</td>
<td>0.521988</td>
</tr>
<tr>
<td>$V(\beta_1^*)$</td>
<td>0.0168358</td>
<td>0.0128648</td>
<td>0.011627</td>
<td>0.0108254</td>
</tr>
<tr>
<td>$V(\beta_1^*)$ % OLS</td>
<td>100.00%</td>
<td>76.4%</td>
<td>69.06%</td>
<td>64.2%</td>
</tr>
<tr>
<td>$V(\beta_1^<em>) + E(\beta_1^</em> - \beta_1^2)$</td>
<td>0.0168358</td>
<td>0.0259050</td>
<td>0.0405679</td>
<td>0.0444289</td>
</tr>
<tr>
<td>% OLS</td>
<td>100.00</td>
<td>153.86</td>
<td>240.96</td>
<td>263.89</td>
</tr>
<tr>
<td>$\beta_2^*$</td>
<td>0.29468</td>
<td>0.05447</td>
<td>-0.038079</td>
<td>-0.0054505</td>
</tr>
<tr>
<td>$V(\beta_2^*)$</td>
<td>0.0534989</td>
<td>0.0256152</td>
<td>0.0172359</td>
<td>0.0158737</td>
</tr>
<tr>
<td>% OLS</td>
<td>100.00</td>
<td>47.879</td>
<td>32.217</td>
<td>29.67</td>
</tr>
<tr>
<td>$V(\beta_2^<em>) + E(\beta_2^</em> - \beta_2^2)$</td>
<td>0.0534989</td>
<td>0.083317</td>
<td>0.127965</td>
<td>0.137804</td>
</tr>
<tr>
<td>% OLS</td>
<td>100.00</td>
<td>153.73</td>
<td>239.13</td>
<td>257.58</td>
</tr>
<tr>
<td>$\beta_3^*$</td>
<td>-1.1690</td>
<td>-0.85999</td>
<td>-0.72840</td>
<td>-0.703438</td>
</tr>
<tr>
<td>$V(\beta_3^*)$</td>
<td>0.065073</td>
<td>0.0296403</td>
<td>0.0191532</td>
<td>0.0174674</td>
</tr>
<tr>
<td>% OLS</td>
<td>100.00</td>
<td>45.549</td>
<td>29.433</td>
<td>26.84</td>
</tr>
<tr>
<td>$V(\beta_3^<em>) + E(\beta_3^</em> - \beta_2^2)$</td>
<td>0.065073</td>
<td>0.125123</td>
<td>0.213281</td>
<td>0.234215</td>
</tr>
<tr>
<td>% OLS</td>
<td>100.00</td>
<td>192.28</td>
<td>327.756</td>
<td>359.9</td>
</tr>
</tbody>
</table>

*MSE ($\beta_1^*$) is estimated using OLS $\beta_1$ as the true $\beta$ since the true $\beta$ are unknown.*
<table>
<thead>
<tr>
<th></th>
<th>OLS K = 0</th>
<th>KA</th>
<th>KB</th>
<th>KM</th>
<th>OLS K = 0</th>
<th>KA</th>
<th>KB</th>
<th>KM</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_1$</td>
<td>0.705293</td>
<td>0.591099</td>
<td>0.533813</td>
<td>0.521988</td>
<td>0.705293</td>
<td>0.591099</td>
<td>0.533813</td>
<td>0.521988</td>
</tr>
<tr>
<td>V($\beta_1$)</td>
<td>0.0168358</td>
<td>0.0128648</td>
<td>0.011627</td>
<td>0.0108254</td>
<td>0.34</td>
<td>0.22</td>
<td>0.21</td>
<td></td>
</tr>
<tr>
<td>V($\beta_1^* - \beta_1^*$)</td>
<td>0.57002</td>
<td>0.0177410</td>
<td>0.0113200</td>
<td>0.0108259</td>
<td>0.76</td>
<td>0.69</td>
<td>0.64</td>
<td></td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>0.29468</td>
<td>0.05447</td>
<td>-0.038079</td>
<td>-0.054505</td>
<td>0.47</td>
<td>0.32</td>
<td>0.29</td>
<td></td>
</tr>
<tr>
<td>V($\beta_2$)</td>
<td>0.05346989</td>
<td>0.0256152</td>
<td>0.0172359</td>
<td>0.0158737</td>
<td>0.43</td>
<td>0.27</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>V($\beta_2^* - \beta_2^*$)</td>
<td>0.534774</td>
<td>0.231303</td>
<td>0.147543</td>
<td>0.0108259</td>
<td>0.43</td>
<td>0.27</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>-0.11690</td>
<td>-0.85996</td>
<td>-0.7284</td>
<td>-0.703438</td>
<td>0.45</td>
<td>0.29</td>
<td>0.26</td>
<td></td>
</tr>
<tr>
<td>V($\beta_3$)</td>
<td>0.065073</td>
<td>0.0296403</td>
<td>0.0191532</td>
<td>0.0174674</td>
<td>0.36</td>
<td>0.19</td>
<td>0.16</td>
<td></td>
</tr>
<tr>
<td>V($\beta_3^* - \beta_3^*$)</td>
<td>0.657880</td>
<td>0.242103</td>
<td>0.127618</td>
<td>0.1101130</td>
<td>0.36</td>
<td>0.19</td>
<td>0.16</td>
<td></td>
</tr>
</tbody>
</table>

*Restriction has been made as to the true coefficient as follows:

$\beta_1^* = 0.52127$
$\beta_2^* = -0.39906$
$\beta_3^* = -0.39906$
(ii) Again the precision of the estimates is recognized and the estimates of $\hat{\beta}_2$ changed from statistically insignificant to statistically significant at a 5% level.

(3) RIDGM Performances

(i) $\hat{\beta}_2$ also changed from 0.2948 insignificant wrong positive sign to -.054505.

(ii) All estimates were possibly statistically significant at a 5% level.

(iii) Gain in precision was obtained since $V(\hat{\beta}_1)$ decreased by about 70% and $V(\hat{\beta}_3)$ decreased by about 73%.

Some Significant Advantages Over OLS

Two important advantages to ridge regression applied over OLS can be noted from Table 10.

(1) The coefficient $\hat{\beta}_2$ of the variable MLS no longer carries a wrong sign under RIDGM and $K_B$. The symptoms of nonsense regression caused by multicollinearity have been corrected.

(2) Precision has been improved under the suggested ridge, and as a result, coefficients are now different from zero at a 5% level, if possible bias effects are ignored.

To conclude this analysis, RIDGM and Lawless-Wang ridge have very important advantages over OLS in improving the estimate in a model which suffers from multicollinearity.
Estimating the Demand Function for Commercially Caught Salmon in the Columbia River

Background

Objectives of the study as stated in Brown et al. (3) can be summarized as follows.

(1) A necessary step in estimating benefits to consumers from commercially caught salmon is to estimate their demand for salmon.

(2) The demand function can be used to compute the prices consumers will be willing to pay for specified quantities of the salmon.

The Demand Function

Basic data is given in Brown et al. (3) (Appendix Tables 1-2) and a demand function is specified in the form

\[ P_{ft} = f(INC_t, PR_t, QF_t) \]

where \( P_{ft} \) denotes the wholesale price of fresh and frozen Chinook salmon in New York for the \( t \)th year, deflated by the wholesale price index; \( INC_t \) is U.S. per capita disposable personal income deleted by the consumer price index; \( PR_t \) is the price of round steak deflated by the consumer price index; \( QF_t \) is U.S. per capita consumption of fresh and frozen salmon.
**Multicollinearity**

Fortunately, multicollinearity is not a problem in the model since the diagonal elements of the inverted correlation matrix (VIF) showed no severe signs of multicollinearity. VIF were

<table>
<thead>
<tr>
<th>Variables</th>
<th>VIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>INC</td>
<td>2.07</td>
</tr>
<tr>
<td>PR&lt;sub&gt;t&lt;/sub&gt;</td>
<td>1.04</td>
</tr>
<tr>
<td>QF&lt;sub&gt;t&lt;/sub&gt;</td>
<td>2.07</td>
</tr>
</tbody>
</table>

The simple correlation matrix shown in Table 13 shows low correlation except for \( r_{13} = -0.76686 \). Other correlations were \( r_{12} = -0.1965 \) and \( r_{23} = 0.18710 \). No harmful multicollinearity is detected in the model.

**Ordinary Least Square Estimates**

This model is a case where ordinary least square performance is expected to be effective and accurate because there is no multicollinearity. As shown in Table 12

1. All coefficients have the correct sign.
2. All variables were highly significant at the 5% level.
3. \( \hat{V}(\beta_1) \) are not unreasonably large.

**Ridge Regressions**

Ridge regression models are not expected to do much better than
Table 12. Coefficients, means and standard deviations for commercially caught salmon in Columbia River demand relationship.

<table>
<thead>
<tr>
<th></th>
<th>INC.</th>
<th>$P_{R_t}$</th>
<th>$Q_{f_t}$</th>
<th>$P_{f_t}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_i$</td>
<td>0.000422</td>
<td>0.21262</td>
<td>-1.45308</td>
<td></td>
</tr>
<tr>
<td>mean $\bar{x}$</td>
<td>2362.21</td>
<td>1.18336</td>
<td>0.199357</td>
<td>0.84878</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>46.0965</td>
<td>0.0851157</td>
<td>0.0504147</td>
<td>0.25827</td>
</tr>
</tbody>
</table>
Table 13. Simple correlation coefficients between variables for commercially caught salmon in Columbia River.

<table>
<thead>
<tr>
<th></th>
<th>$P_f_t$</th>
<th>INC</th>
<th>$P_R_t$</th>
<th>$Q_f_t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_f_t$</td>
<td>1.000000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INC</td>
<td>0.042967</td>
<td>1.000000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$P_R_t$</td>
<td>-0.1310547</td>
<td>-0.196504</td>
<td>1.000000</td>
<td></td>
</tr>
<tr>
<td>$Q_f_t$</td>
<td>0.810522</td>
<td>-0.716686</td>
<td>0.187101</td>
<td>1.000000</td>
</tr>
</tbody>
</table>
Table 14. Demand relationships for the commercially caught salmon in Columbia River coefficients, variances and estimated mean square errors for OLS, $K_A$, $K_B$ and $K_M$ estimators.

<table>
<thead>
<tr>
<th></th>
<th>OLS $K = 0$</th>
<th>$K_B = K_M$</th>
<th>$K_A$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_1*$</td>
<td>0.000422</td>
<td>0.000409</td>
<td>0.000404</td>
</tr>
<tr>
<td>$V(\beta_1^*)$</td>
<td>0.0059436</td>
<td>0.0052737</td>
<td>0.0060166</td>
</tr>
<tr>
<td>$V(\beta_1^<em>) + E(\beta_1^</em> - \beta_1^*)^2$</td>
<td>0.0059436</td>
<td>0.005755</td>
<td>0.0049847</td>
</tr>
<tr>
<td>$\beta_2*$</td>
<td>0.21262</td>
<td>0.200708</td>
<td>0.195019</td>
</tr>
<tr>
<td>$V(\beta_2^*)$</td>
<td>0.002995</td>
<td>0.00287127</td>
<td>0.0028120</td>
</tr>
<tr>
<td>$V(\beta_2^<em>) + E(\beta_2^</em> - \beta_2^*)^2$</td>
<td>0.002995</td>
<td>0.002886</td>
<td>0.0028456</td>
</tr>
<tr>
<td>$\beta_3*$</td>
<td>-1.45308</td>
<td>-1.4999</td>
<td>-1.51986</td>
</tr>
<tr>
<td>$V(\beta_3^*)$</td>
<td>0.005921</td>
<td>0.0052556</td>
<td>0.004968</td>
</tr>
<tr>
<td>$V(\beta_3^<em>) + E(\beta_3^</em> - \beta_3^*)^2$</td>
<td>0.005921</td>
<td>0.0053393</td>
<td>0.0051383</td>
</tr>
</tbody>
</table>
ordinary least square models because of the fact that the data is close to Orthogonality and only a small improvement can be expected.

Comparison Between OLS and Ridge Regression Selected Models

As shown from Table 14 all ridge models \( \mathbf{K}_A, \mathbf{K}_B, \mathbf{K}_M \) perform as well as ordinary least square models with ridge regression carrying a slight advantage over OLS in terms of

1. \( \hat{V}(\hat{\beta}_i) \) slightly reduced in ridge models

2. The measurement of MSE(\( \hat{\beta}_i \)) in terms of MSE(\( \hat{\beta}_i \)) = \( \hat{V}(\hat{\beta}_i) + E(\hat{\beta}_i - \hat{\beta}_i)^2 \) indicating that ridge models show slight improvement over OLS.

As stated earlier, it is not surprising that OLS performs so well because of the nature of \( \mathbf{x} \) data that tends toward Orthogonality. Unfortunately economists do not always have access to such high quality data.
V. SUMMARY AND CONCLUSIONS

When there is a relatively severe multicollinearity problem in economic data, $x'x$ will be ill-conditioned\(^{12}\) and ordinary least square estimates will be both unstable and unreliable. The variances of the OLS coefficients will be unreasonably high, and $t$ and $F$ tests will often result in the rejection of more important variables considered statistically the same as zero at any reasonable significance level.

In an attempt to solve the problem, many methods have been suggested and practiced. Some require that data be collected outside the multicollinearity range, while for others, variables must be either combined or deleted from the economic model.

Frequently, these practices result in inadequate prediction equations giving poor and sometimes erroneous information about the population under investigation. Yet at the same time, there are more promising techniques demanding more information for the improvement of regression estimates in the presence of multicollinearity. The method of Theil and Goldberger (14) is one way of incorporating prior information into the regression problem to lessen the difficulties created by multicollinearity. But this method, also, is not without drawbacks caused by dominance of the prior information over

\(^{12}\) Generally speaking, $(x'x)$ is ill conditioned when one or more of the eigenvalues of $(x'x)$ are close to zero.
the sample information, and sometimes because of the biased nature of the prior information.

Ridge regression can be considered as a method of incorporating prior information into the regression analysis to solve the multicollinearity problem, where the prior $\beta$ vector is the null vector. Ridge is among a class of biased linear estimators. It yields regression coefficients that are biased in terms of the true $\beta$ vector. But these coefficients have a very desirable property, which is the relatively small variance associated with the estimates. The gain in precision obtained by the use of ridge regression overweighs the bias introduced to the system, as measured by $\text{MSE}(\hat{\beta})$. The bias is much smaller than the variances of ordinary least square as indicated by the Monte Carlo studies by Hoerl, Kennard and Baldwin (11) and by Lawless and Wang (16). RIDGM, a version of ridge regression, is also a promising method for improving estimates highly affected by multicollinearity.

When RIDGM was compared to ordinary least square, it appeared to have important advantages.

1. RIDGM estimates were biased but theoretically they were more stable and reliable, while ordinary least square yielded estimates of the wrong sign and offered no significant meaning when interpreted. RIDGM yielded statistically significant estimates of the right sign, if possible bias is ignored.
(2) The variances of the coefficients obtained by RIDGM were relatively smaller compared to those obtained by ordinary least square.

The other two versions of ridge regression, $K_A$ and $K_B$ also yielded estimates with variances much smaller than that obtained by OLS.

In conclusion, significant improvement is not expected from the use of ridge regression over ordinary least squares, when the data are not ill-conditioned or near Orthogonality. In such cases, ordinary least square will yield estimates that are unbiased with variances not much larger than for ridge regression.

Limitations and Additional Research Needed

Although ridge regression performed well for the empirical models considered, additional questions raised by this research are the following:

(1) Which method for estimating the biasing parameter $K$ is best? For the models considered, RIDGM appeared to do well. However, an extensive Monte Carlo study would be needed, where RIDGM and the other methods for selecting $K$ could be compared.

(2) How would ridge regression compare with OLS when there are other defects in the data, in addition to multicollinearity? For example, errors in measurement of the explanatory variables are
known to result in bias when OLS is used. Would such a problem be made worse or better when ridge regression is employed? Again, an extensive Monte Carlo experiment under carefully specified conditions would seem to be a promising method for answering such questions.

In summary, much additional research is still needed to adequately evaluate the role of biased linear estimation in economic research. However, for the four economic models considered in this thesis, the use of ridge regression appears to be very promising given the recent advances in the selection of the biasing parameter, $K$. 
BIBLIOGRAPHY


